

Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward

Year: 2021 | Citations: 392 | Authors: Momina Masood, M. Nawaz, K. Malik, A. Javed, Aun Irtaza

Abstract

Easy access to audio-visual content on social media, combined with the availability of modern tools such as Tensorflow or Keras, and open-source trained models, along with economical computing infrastructure, and the rapid evolution of deep-learning (DL) methods have heralded a new and frightening trend. Particularly, the advent of easily available and ready to use Generative Adversarial Networks (GANs), have made it possible to generate deepfakes media partially or completely fabricated with the intent to deceive to disseminate disinformation and revenge porn, to perpetrate financial frauds and other hoaxes, and to disrupt government functioning. Existing surveys have mainly focused on the detection of deepfake images and videos; this paper provides a comprehensive review and detailed analysis of existing tools and machine learning (ML) based approaches for deepfake generation, and the methodologies used to detect such manipulations in both audio and video. For each category of deepfake, we discuss information related to manipulation approaches, current public datasets, and key standards for the evaluation of the performance of deepfake detection techniques, along with their results. Additionally, we also discuss open challenges and enumerate future directions to guide researchers on issues which need to be considered in order to improve the domains of both deepfake generation and detection. This work is expected to assist readers in understanding how deepfakes are created and detected, along with their current limitations and where future research may lead.